



UNIVERSITÉ DE NANTES



IAE NANTES
ÉCONOMIE & MANAGEMENT

ANALYSE DES DONNÉES DE PANEL

Analyse du poids des bébés à la naissance



Teodoro MOUNIER TEBAS
Diane THIERRY

Enseignant : M. WOLF
Année universitaire : 2020-2021

Table des matières

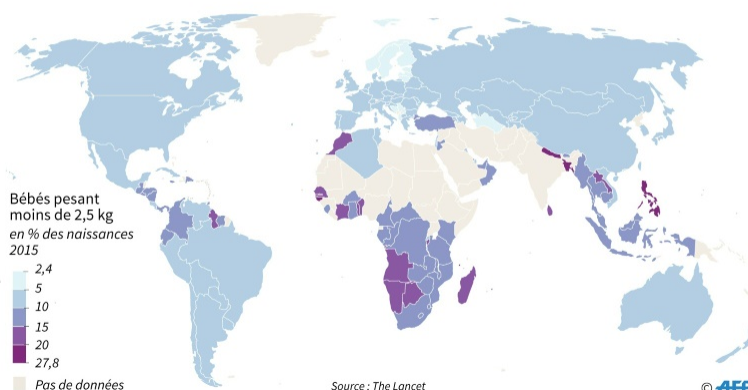
1	Introduction	2
2	Description des données	3
2.1	Variable à expliquer	3
2.2	Variables explicatives	4
3	Statistiques descriptives des variables sélectionnées	9
4	Analyse économétrique	11
4.1	Méthodologie	11
4.2	Modélisations	13
4.2.1	Détection de l'hétérogénéité	13
4.2.2	Test d'homogénéité totale	13
4.2.3	Estimations	14
5	Conclusion	15

1 Introduction

Chaque année, plus de 20 millions de nouveau-nés, dont plus de 96% dans les pays en développement, pèsent moins de 2500 g à la naissance d'après l'OMS. Parmi les 2,5 millions de bébés qui meurent à la naissance, plus de 80% souffrent d'insuffisance pondérale qui accroît le risque de morbidité et de mortalité néonatales.¹ Tandis que le poids normal d'un nouveau-né se situe entre 2,5 et 4,3 kg, ceux dont le poids est inférieur à 2,5 kg sont considérés de **faible poids**. Cette situation augmente le risque de décès prématuré ainsi que de pathologies chroniques plus tard dans la vie, telles que troubles cardiovasculaires, croissance inférieure à la moyenne, problèmes respiratoires, problèmes neurologiques, diabète, cécité, surdité, problèmes de comportement et difficultés d'apprentissage...²

20,5 millions de bébés avec un faible poids de naissance

Soit 14,6% des naissances dans le monde en 2015



La naissance de bébés avec un faible poids est une vérité qui n'épargne aucun pays même si elle semble moins impacter les pays développés, comme nous pouvons le voir sur la carte ci-dessus. De nombreuses études ont effectivement montré qu'une cause importante d'insuffisance pondérale est le cadre socio-économique dans lequel le bébé naît. Dans les pays développés les faibles poids sont plus souvent liés à des naissances prématurées, tandis que dans les pays moins développés cela peut être lié à une prévalence élevée d'infections, des complications de grossesse ou des conditions de vie défavorisées de manière générale. Dans tous les cas, des travaux ont montré comment l'environnement social (le stress chronique lié au logement, aux violences conjugales, aux problèmes financiers etc.), l'hygiène de vie (alimentation, tabagisme, alcoolisme) et les caractéristiques de la mère (taille, poids, maladies) influencent le poids des bébés à la naissance.

Cependant ces conditions de vie diffèrent grandement d'un continent à un autre, voire au sein d'un même pays, définissant alors la santé des nouveau-nés dès leur conception. Dans cette étude nous tenterons de répondre à la problématique suivante : "Quels facteurs influencent le taux d'insuffisance pondérale dans les pays développés ?". Nous centrons ainsi notre analyse sur les proches voisins de la France, ainsi que sur les années allant de 2000 à 2015.

Dans une première partie nous justifierons et expliquerons chacune des variables que nous avons sélectionnées pour l'analyse, variables que nous examinerons au travers de statistiques descriptives dans une deuxième partie. Puis, nous terminerons ce dossier avec l'application de modèles économétriques visant à modéliser les relations entre les variables, et à expliquer les variations de Y.

1. <https://news.un.org> (consulté le 26 janvier)

2. <https://www.inspq.qc.ca> (consulté le 26 janvier)

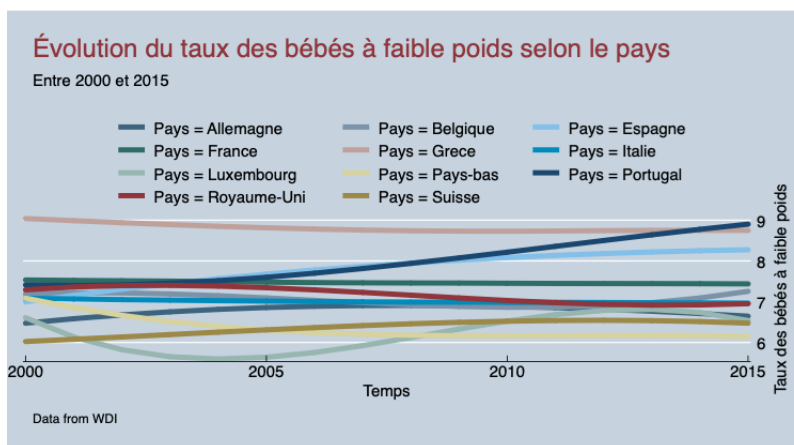
2 Description des données

2.1 Variable à expliquer

Comme présenté en introduction, nous allons étudier le poids des bébés à la naissance et plus précisément le nombre de nouveau-nés de faible poids, en pourcentage des naissances. Il faut tout d'abord savoir que cette situation dangereuse pour la santé des bébés n'est peut être pas pleinement mesurée puisque près d'un tiers des nouveau-nés ne sont pas pesés à la naissance.

La mesure du poids à la naissance est prise dans la première heure de vie du bébé, avant qu'une perte de poids postnatale significative ne se produise. Les premiers jours de sa vie, le nouveau-né peut effectivement perdre environ 10% du poids qu'il avait à la naissance. Cette situation est tout à fait normale et il retrouvera son poids de naissance entre 10 et 14 jours. En revanche les bébés nés prématurément, soit avant la 37^e semaine de grossesse, peuvent être plus petits de poids et de taille que les autres nouveau-nés, et leur rythme de croissance pourrait aussi être plus lent au début.³

Nous avons décidé de cibler notre analyse sur les pays européens car les causes de faible poids des bébés sont moins évidentes que dans des régions en voie de développement où le cadre sanitaire et financier joue en grande partie. Nous avons donc choisi les pays frontaliers (ou presque) à la France, tels que : Allemagne, Pays-Bas, Belgique, Luxembourg, Royaume-Uni, Suisse, Italie, Grèce, Espagne et Portugal. De même, nous avons voulu expliquer ce phénomène sur des données relativement récentes pour lesquelles nous avons le plus d'information soit de 2000 à 2015.



Le graphique ci-dessus nous montre l'évolution du pourcentage des nouveau-nés avec un poids faible entre 2000 et 2015, selon le pays considéré. On voit alors que le taux est le plus élevé pour la Grèce dont les valeurs sont assez constantes sur cette période, tandis que le pays où cette proportion est plus faible est le Luxembourg entre 2001 et 2008, puis les Pays-Bas à partir de la crise financière mondiale. Le taux de bébés à faible poids en France semble constant sur la période, variant autour de 7.5% des naissances. Enfin, un dernier constat intéressant est celui de la hausse constante du chiffre au Portugal depuis les années 2000 ; allant de 7.4 à 8.9% en 15 ans.

Comme l'a précisé la Dr Mercedes de Onis de l'OMS en mai dernier ; « La réduction de l'insuffisance pondérale à la naissance exige la compréhension des causes sous-jacentes dans un pays donné ». ⁴ Voyons quels facteurs nous avons sélectionnés pour expliquer les variations de Y dans le temps et dans l'espace.

3. <https://naitreetgrandir.com> (consulté le 09/11/2020)

4. <https://news.un.org> (consulté le 27/01/2021)

2.2 Variables explicatives

TABLE 1 – Variables sélectionnées pour expliquer le faible poids à la naissance

Caractéristiques de l'environnement social	- Emploi à temps partiel, femmes (% de l'emploi total chez les femmes) - Chômage (% de la population active)
Caractéristiques de la mère	- Prévalence tabagisme (% des adultes) - Prévalence de l'anémie chez les femmes enceintes (%)
Caractéristiques du fœtus	- Taux de fécondité (nb de naissance par femme) - Ratio des sexes à la naissance (filles pour 1,000 garçons)

À partir des différentes recherches bibliographiques et au vu des variables disponibles dans la base de données *World Development Indicators* (WDI) disponible en open data sur le site de la Banque Mondiale, nous avons sélectionné 6 facteurs explicatifs pouvant être regroupés en 3 catégories. Les **caractéristiques de l'environnement social** que nous présentons en introduction et qui définissent l'état psychologique et la santé de la mère et donc celle de son futur bébé. Une variable concernant la situation de la mère (mariée, seule, en couple) aurait été intéressante mais il s'agirait alors de données micro-économiques non disponibles sur WDI.

La deuxième catégorie est celle des **caractéristiques de la mère** qui sont censées refléter son hygiène de vie, et la troisième concerne les **caractéristiques de l'enfant**. Certaines études ont effectivement montré plusieurs tendances : les premiers nés, les filles et les jumeaux (triplets ou plus) ont tendance à avoir un poids plus léger et donc un risque plus grand d'insuffisance pondérale à la naissance.

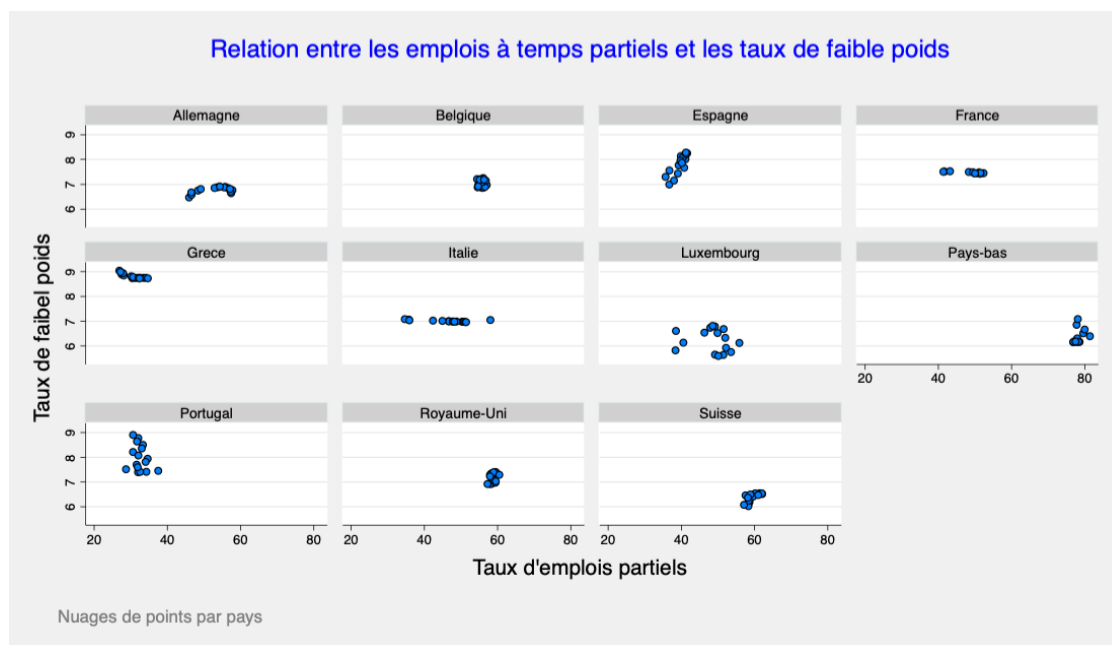
A- Taux d'emplois à temps partiel chez les femmes

Le **temps partiel** fait référence à un emploi régulier dans lequel le temps de travail est nettement inférieur à la normale. Lorsqu'ils ne sont pas choisis, les emplois à temps partiel mettent dans une situation d'instabilité et conduisent à la précarité et à la hausse des inégalités. En effet, malgré une forme d'emploi stable (les 3/4 sont en CDI), le temps partiel se caractérise par le fait que les salariés sont souvent très peu qualifiés et on retrouve donc beaucoup de smicards et de bas salaires.⁵

De plus, plus de 80% des emplois partiels sont occupés par des femmes, accentuant ainsi les inégalités hommes-femmes et mettant ces dernières dans une situation précaire, davantage encore si elles ont un enfant. Nous nous attendons alors à une **relation positive** entre cette variable et Y puisqu'une proportion élevée d'emplois à temps partiel augmente le stress chronique de la mère et accroît les chances d'insuffisance pondérale pour l'enfant à venir.

Cependant, sur notre échantillon cette hypothèse n'est que très peu vérifiée puisque comme visible sur le graphique suivant, les types de relations sont difficilement identifiables, excepté pour l'Allemagne et l'Espagne où on voit clairement une relation positive. On voit aussi que le taux d'emplois à temps partiel est plus faible dans les pays du sud tels que le Portugal, l'Espagne et la Grèce.

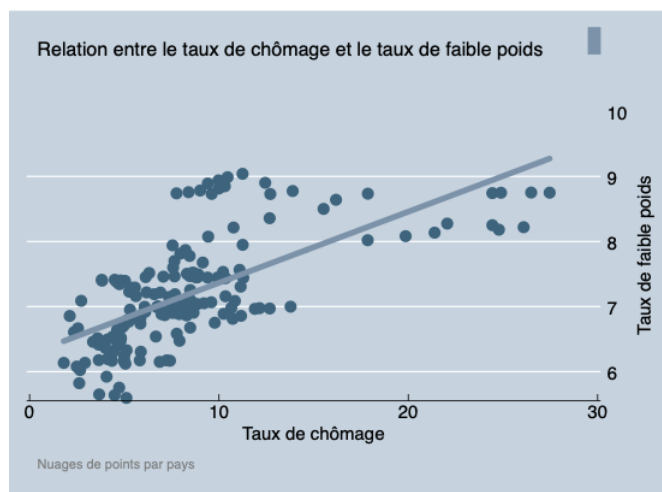
5. <https://www.focusrh.com> consulté le 27/01



B- Taux de chômage

Selon l'INSEE, le **taux de chômage** correspond au “pourcentage de chômeurs dans la population active” ; un chômeur est donc une personne n’ayant pas de travail mais en cherchant activement un. Comme le taux d’emplois à temps partiel, cette variable a pour but de refléter le contexte macro-économique de la période et du pays étudié, qui peut alors impacter sur la qualité de vie de la mère et donc celle de son enfant.

De manière générale le taux de chômage est une variable qui connaît beaucoup de fluctuations : les périodes de récession se traduisent souvent par une hausse des licenciements, donc du nombre de chômeurs, alors que les périodes d’expansion sont synonymes de création d’emplois et de baisse du taux de chômage. Sachant que notre période, qui englobe les 15 premières années du siècle, passe par la crise des Subprimes en 2008, nous nous attendons évidemment à une hausse du taux de chômage, en espérant que celui-ci ne bouscule pas trop les relations avec la variable à expliquer.



On identifie cette fois clairement une relation positive entre le taux de chômage et le taux de bébés nés avec un poids faible ; toutes années et tous pays confondus, l’hypothèse est donc validée.

C- Prévalence du tabagisme

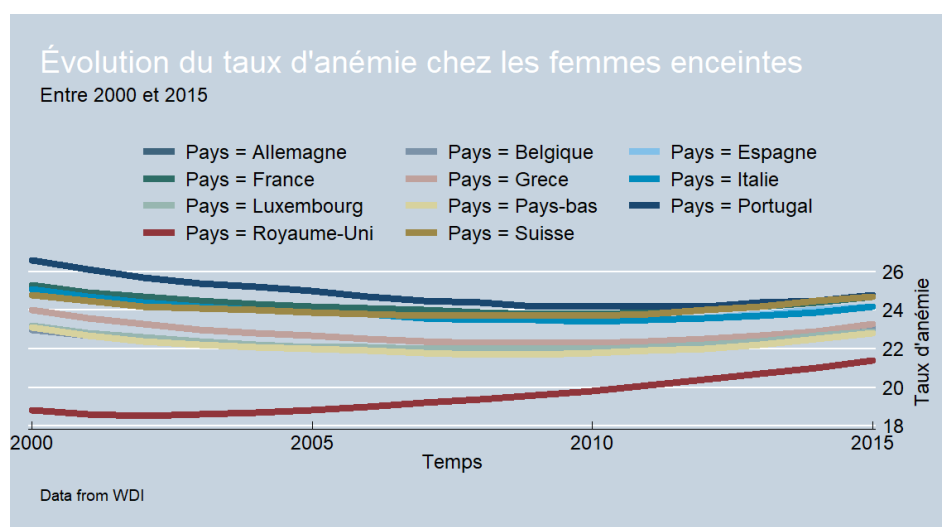
Cette variable récoltée aussi dans la base WDI, correspond au pourcentage de la population âgée de 15 ans et plus qui **consomme actuellement un produit du tabac** quotidiennement ou non. Cela comprend les cigarettes, les pipes, les cigares, les cigarillos, les pipes à eau (narguilé, chicha), les bidis, le kretek, les produits du tabac chauffés et toutes les formes de tabac sans fumée (oral et nasal). Nombre d'études et d'articles parlent de l'importance de l'hygiène de vie de la mère pour la santé de son enfant. La consommation d'alcool ou de tabac peut détériorer rapidement l'état de ce dernier et peut ainsi accroître les risques d'insuffisance pondérale à la naissance.⁶

Cette information va donc de pair avec la qualité de l'alimentation, la consommation de drogues et le diabète. Nous ne pouvions cependant considérer autant de facteurs explicatifs et avons donc choisi la consommation de tabac qui semble relativement élevée et qui peut l'être davantage dans certains pays ou à certaines périodes. La relation que nous attendons serait alors **positive** : un fort taux de tabagisme provoque une santé détériorée (sans oublier le tabagisme passif) et ainsi plus de maladies chez le nouveau-né, dont un poids faible.

D- Prévalence de l'anémie chez les femmes enceintes (%)

La variable "prévalence de l'anémie chez les femmes enceintes" correspond au pourcentage de femmes enceintes dont le taux d'hémoglobine est inférieur à 110 grammes par litre de sang selon la **World Health Organisation**⁷. L'anémie n'est pas vraiment une maladie mais plutôt un symptôme du à la difficulté du sang à transporter l'oxygène, nécessaire à la production d'énergie des cellules. L'anémie provoque un teint pâle, de la fatigue et une sensation de froid.

Lorsque nous réduisons la population aux femmes enceintes, la principale cause de l'anémie est une carence en fer qui peut venir par une perte récurrente de fer au cours des règles ce que l'on appelle l'**anémie ferriprive**. En plus des complications telles que fatigue, vertiges, dyspnée d'effort modérée, l'anémie peut aussi être à l'origine d'accouchements prématurés.⁸ L'accouchement prématuré étant bien entendu l'une des causes du poids trop faible des bébés à la naissance.



6. <https://www.inspq.qc.ca>

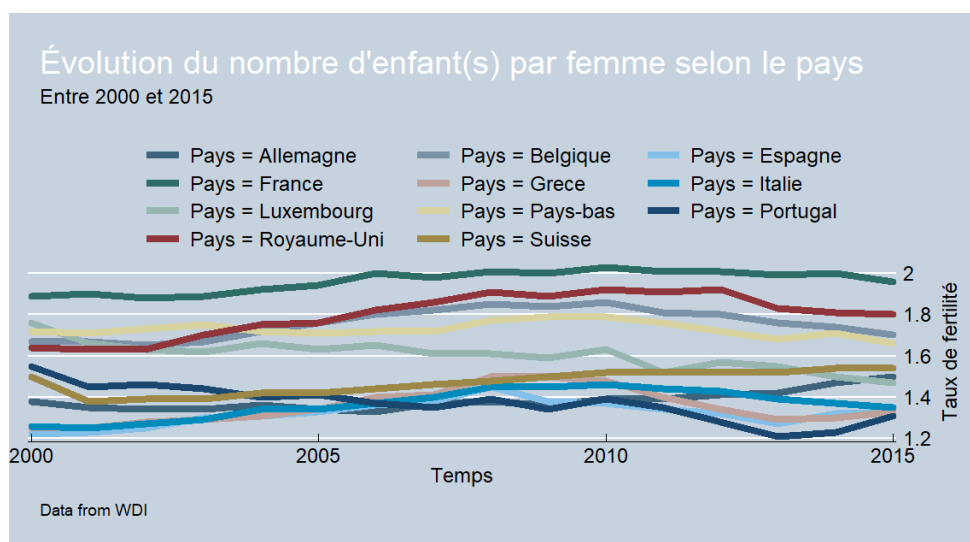
7. Organisation créée en 1948 travaillant avec 194 États membres, dans six régions et dans plus de 150 bureaux, le personnel de l'OMS est uni dans un engagement commun visant à améliorer la santé pour tous, partout.

8. Un travail (contractions résultant des modifications du col utérin) qui débute avant 37 semaines est considéré comme prématuré.

Comme nous pouvons le constater sur le graphique précédent, c'est au Royaume-Uni que le pourcentage des femmes enceintes anémiques est le plus faible (22% en 2015) et ce, tout au long de la période étudiée, c'est-à-dire depuis les années 2000 jusqu'à 2015. D'une manière générale, sur ces 15 années et quelque soit le pays, les taux sont compris entre 22% et 26% des femmes enceintes. C'est un taux assez important ce qui en fait une variable non-négligeable dans l'explication du faible poids des bébés à la naissance.

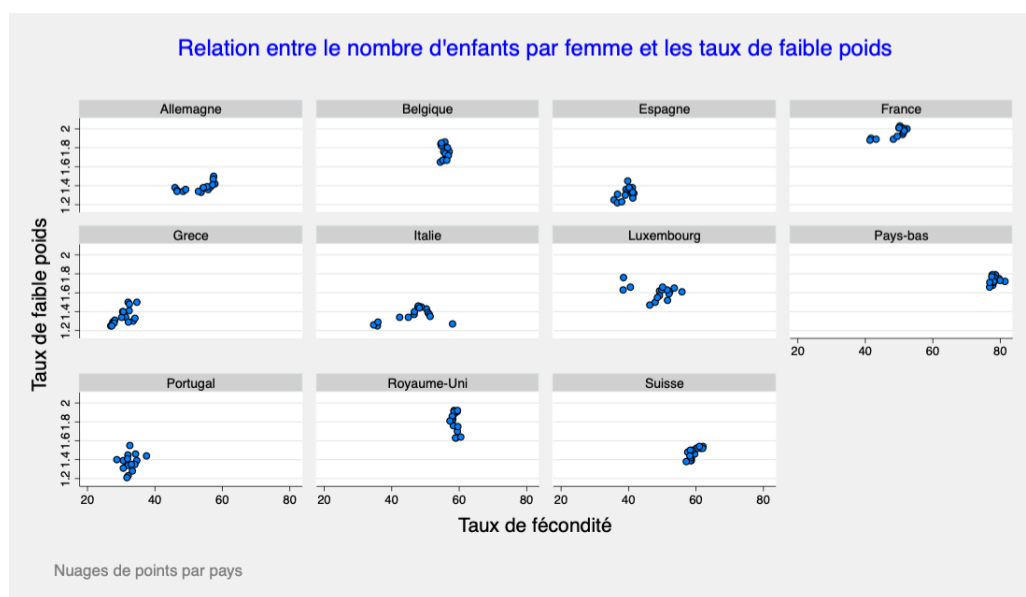
E- Taux de fécondité

Le **Taux de fécondité** représente le nombre d'enfant(s) par femme si celle-ci devait vivre jusqu'à la fin de ses années de procréation et avoir des enfants conformément. C'est un taux qui est très utilisé lorsqu'il s'agit d'étudier la reproduction sociale ainsi que notre système de retraites (notamment français) par répartition⁹. En effet, les faibles taux de fertilité européens ne permettent pas le renouvellement de la population ce qui compromettra l'équilibre des systèmes de retraite fondés sur une solidarité entre les générations. Dans notre étude, le taux de fertilité pourrait expliquer pourquoi certains pays font face à des naissances où les bébés ont un poids trop faible. L'hypothèse étant la suivante ; plus une femme a d'enfants, plus la probabilité que celui-ci naisse avec un poids trop faible augmente. Nous verrons si cela se confirme à travers les modélisations. Nous allons tout de même observer ce taux sur la période de 2000 à 2015 pour les 11 pays européens de notre étude.



Comme nous pouvions nous en douter, la courbe française est bien la plus élevée avec une moyenne de deux enfants par femmes ce qui est supérieur à nos voisins européens. Nous devrions donc constater, à partir de notre hypothèse de départ, un nombre de naissances de bébés au poids faible plus grand, proportionnellement aux autres pays. Nous pouvons observer cela à partir de la figure suivante représentant la variable "taux de fécondité" relativement à notre variable à expliquer.

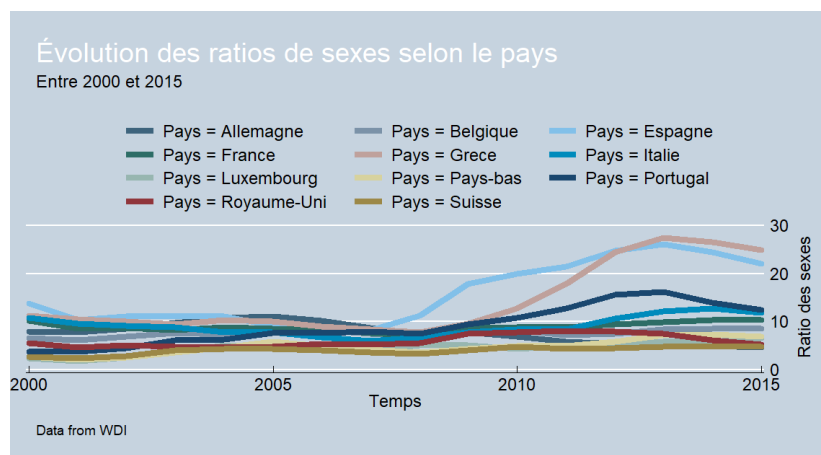
9. Les cotisations des actifs versées aujourd'hui servent à payer immédiatement les pensions versées aux retraités.



On observe donc bien un taux supérieur de bébés avec un trop faible poids à la naissance pour la France. La Belgique et le Royaume-Uni suivent la France de très près. Les portugais ainsi que les suisses ou encore les espagnols font face à des taux très faibles d'enfants naissants avec un faible poids.

F- Ratio des sexes à la naissance (filles pour 1,000 garçons)

Le **ratio des sexes à la naissance** est la dernière variable explicative choisie dans l'explication de notre variable endogène. En effet, pour le même âge gestationnel, les filles pèsent moins que les garçons. Le ratio des sexes à la naissance faisant référence aux naissances masculines par naissance, cette variable devrait se retrouver significative dans l'explication des bébés à faible poids à la naissance. Les données sont des moyennes sur 5 ans, observons celles-ci sur la période de 2000 à 2015 à travers le graphique suivant :



Nous pouvons constater de manière surprenante que le nombre de naissances de bébés garçons est beaucoup plus faible que celui de petites filles, avec toutefois un taux plus élevé pour l'Espagne ainsi que la Grèce entre 2007 et 2015. Maintenant que les variables explicatives et à expliquer ont été présentées, nous pouvons passer au traitement de la base et statistiques descriptives nécessaires avant les modélisations.

3 Statistiques descriptives des variables sélectionnées

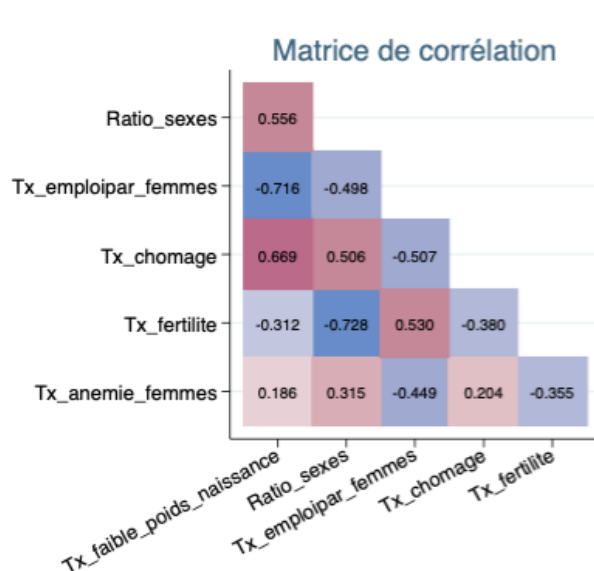
L'étude que nous réalisons s'appuie sur une base de données comportant les 7 variables définies en partie précédente, pour 11 pays européens et sur 15 années : avec un total de 176 observations.

FIGURE 1 – Nombre de valeurs manquantes par variable

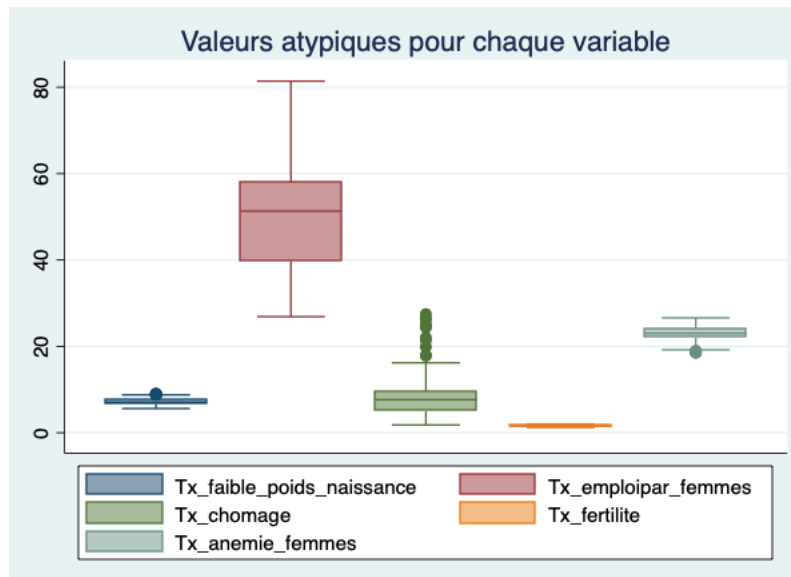
Variable	Obs<.			Obs<.		
	Obs=.	Obs>.	Obs<.	Unique values	Min	Max
Pays	(string variable)					
Annee			176	16	2000	2015
Tx_faible_~e			176	176	5.59675	9.039376
Ratio_sexes	66		110	20	1.046	1.068
Tx_emploi~s			176	168	26.88	81.4
Tx_chomage			176	176	1.805	27.4662
Tx_tabac	132		44	34	22	45.3
Tx_fertilite			176	69	1.21	2.03
Tx_anemie_~s			176	54	18.5	26.6

Comme visible sur la table ci-dessus, au téléchargement de la base depuis le site internet de la Banque Mondiale, nous avons initialement 66 valeurs manquantes pour le ratio des sexes à la naissance, et 132 pour le taux de consommation de tabac. Le ratio des sexes à la naissance n'évolue pas grandement d'année en année, c'est pourquoi nous avons choisi d'imputer la valeur moyenne pour chaque pays du ratio aux NAs. Par exemple, pour le Luxembourg nous avons 10 observations sur 15, nous avons alors calculé la moyenne à partir des 10 existantes, que nous avons imputée aux 5 manquantes. Ces manipulations ont été réalisées directement sous excel.

Pour la variable *Tx_tabac* en revanche, le nombre de valeurs manquantes est plus grand que les données existantes (4 pour 11) et on observe une tendance à la baisse du taux pour chacun des pays, donc imputer une moyenne ou une quelconque autre valeur fixe n'aurait pas été pertinent. Puisque nous n'avons pas assez de données connues nous décidons de retirer cette variable ; nous continuons alors l'analyse avec 5 variables explicatives.



La matrice de corrélation ci-dessus nous montre que les relations entre les variables sont assez importantes : les coefficients de corrélation (CC) vont de 0.186 à 0.728 en valeurs absolues. Les relations les plus fortes sont entre le ratio des sexes et le taux de fécondité ce qui est logique, et entre le temps d'emplois à temps partiel pour les femmes avec Y, c'est-à-dire la proportion de nouveau-nés avec un poids inférieur à 2.5kg. L'application des MCO n'est possible que lorsque les variables explicatives ne sont pas fortement corrélées entre elles, c'est pourquoi nous décidons de retirer le ratio des sexes et continuons donc l'analyse avec 4 X_i .



La figure ci-dessus nous montre cette fois les points atypiques de chaque variable, lorsque l'on considère les 15 années et les 11 pays ensemble. Une valeur est dite atypique lorsqu'elle est supérieure ou inférieure aux 1 ou 3^e quantiles signifiés par les "moustaches" de la boîte. Comme nous l'avons souligné en partie économique, le taux de chômage comporte de nombreuses valeurs atypiques dues à la crise mondiale de 2008 faisant exploser les taux. Quelques observations du taux de faible poids à la naissance (Y) sont elles aussi atypiques ; il s'agit de la Grèce en 2000 où le taux atteint 9.039% tandis que les années après il avoisine les 8.7%. C'est un des pays avec cette proportion la plus grande. Enfin, c'est au Royaume-Uni que le taux d'anémie chez les femmes enceintes est le plus faible, constituant visuellement des points atypiques (autour de 19% des femmes, contre 22.97% en moyenne sur tout l'échantillon). Nous laissons ces points qui malgré leur valeur forte ou faible, contiennent des informations non négligeables pour la suite de l'analyse.

Enfin, quelques statistiques des variables, obtenues par la commande *summarize* nous montrent que les variables avec l'étendue la plus forte sont le taux d'emplois partiels chez les femmes, qui va de 26.88 à 81.4%, ainsi que le taux de chômage (de 1.805 à 27.46%) comme souligné précédemment. Ce taux à 1.805% correspond à la proportion de chômeurs au Luxembourg dans l'année 2001. On voit aussi qu'entre 2000 et 2015 et sur ces quelques pays européens, une femme sur deux travaille à temps partiel (moyenne de la variable de 50.15%). La variable à expliquer a une étendue de moins de 5 points ; allant de 5.6 à 9.04% respectivement pour le Luxembourg en 2004 et la Grèce en 2000.

4 Analyse économétrique

4.1 Méthodologie

Les **données de panel** présentent une double dimension : individuelle et temporelle, qui peut se modéliser de différentes manières en fonction de celle-ci. La spécification du modèle linéaire simple peut s'écrire pour N individus ($i = 1, \dots, N$) et T observations temporelles ($t = 1, \dots, T$), soit $n = N * T$ observations totales, de la manière suivante :

$$y_{it} = \alpha_{0i} + \beta'_i X_{it} + \epsilon_{it}$$

avec :

- y_{it} : variable endogène observée pour l'individu i à la période t
- X_{it} : vecteur de k variables exogènes
- α_{0i} : terme constant pour l'individu i
- β'_i : vecteur des k coefficients des k variables exogènes
- ϵ_{it} : terme d'erreur

À partir de la spécification générale présentée ci-dessus quatre possibilités sont envisageables. Si nous constatons que les constantes α_{0i} et les coefficients β'_i sont tous statistiquement identiques pour tous les individus alors il y aurait une **homogénéité totale** et nous pourrions estimer une seule équation par MCO (ou MCG selon la structure de la matrice des variances-covariances des erreurs). À l'inverse, il y aurait **hétérogénéité totale** si les constantes α_{0i} ainsi que les coefficients β'_i étaient tous différents pour toutes les valeurs de i . Dans ce cas là il faudrait estimer N équations c'est-à-dire une équation par individu (la structure des données de panel serait dans ce cas rejeté). Nous pourrions aussi faire face à une **hétérogénéité des coefficients et une homogénéité des termes constants**. Dans cette situation, comme la précédente, il faudrait estimer par les MCO (ou MCG) N équations. Enfin le dernier cas possible serait de faire face à une **hétérogénéité des termes constants et une homogénéité des coefficients des variables explicatives**. Dans ce cas nous ferions appel aux **modèles à effet individuel**. Afin de déterminer la structure de données de panel, nous allons procéder à des tests développés par Hsiao en 1986 qui seront construits à partir de la statistique de Fisher (test de Wald). Ces tests nous permettront de trouver la meilleure méthode à utiliser pour notre sujet d'étude.

Les modèles à effets individuels sont utilisés lorsque les termes constants sont hétérogènes mais qu'il existe une certaine homogénéité dans les valeurs des coefficients des variables explicatives. Nous pouvons spécifier de deux manières cet effet individuel : soit par des modèles à **effets fixes** où on fait l'hypothèse que l'effet individuel est constant dans le temps, soit par des modèles à **effets aléatoires** où le terme constant est une variable aléatoire. En ce qui concerne les modèles à effets fixe l'**estimateur Within** aussi appelé *estimateur intra-individuel* consiste à centrer toutes les variables à expliquer et explicatives sur leur moyenne individuelle, et à appliquer la méthode des MCO (ou MCG si les erreurs sont hétéroscédastiques ou autocorrélées) ce qui peut s'écrire de la façon suivante :

$$(y_{it} - \bar{y}_i) = \beta'(X_{it} - \bar{X}_i) + \epsilon_{it}$$

Pour $i = 1, \dots, N$ et $t = 1, \dots, T$.

En ce qui concerne les modèles à effets aléatoires, ils supposent une relation aléatoire et non plus fixe entre la variable à expliquer et les variables explicatives. L'effet individuel n'est plus un paramètre fixe α_{0i} mais une variable aléatoire. Le terme d'erreur se décompose de la manière suivante :

$$\epsilon_{it} = \alpha_{0i} + \lambda_t + v_{it}$$

avec :

- α_{0i} : représentent les effets individuels aléatoires
- λ_t : représentent les effets temporels identiques pour tous les individus
- v_{it} : terme d'erreur qui est orthogonal aux effets individuels et temporels

Ce type de modèle est aussi connu sous le nom de modèle à erreur composée où des hypothèses sont faites sur les variables aléatoires. Les espérances nulles des α_{0i} , l'homoscédasticité des λ_t et enfin l'indépendance temporelle entre les trois composantes. La méthode d'estimation appropriée pour les modèles à effets aléatoires est celle des MCG car la composante α_{0i} est présente dans ϵ_{it} ce qui provoque une autocorrélation des erreurs. De plus l'estimateur MCG est au final une moyenne pondérée des estimateurs *Within* et *Between*. L'estimateur **between** (estimateur *inter-individuel*) étant l'estimateur des MCO appliqué sur les données, correspond aux N moyennes individuelles des variables explicatives et à expliquer :

$$\overline{y_i} = \alpha_0 + \alpha_{0i} + \beta' \overline{X_i} + \overline{v_i}$$

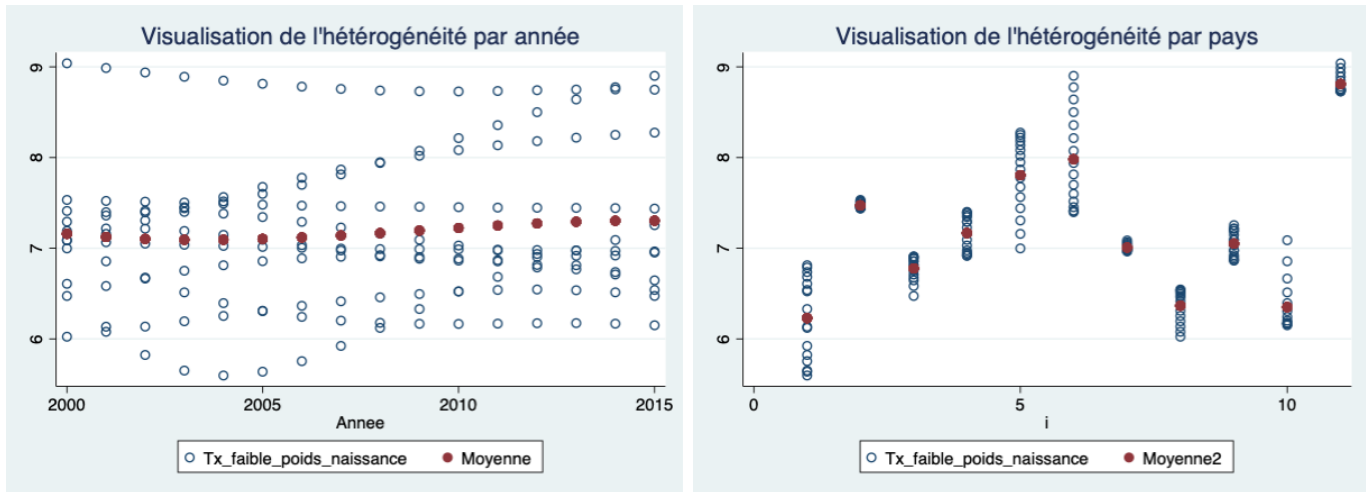
Pour résumer, l'introduction d'effets individuels aléatoires permet de combiner une spécification intermédiaire entre le modèle sans effet individuel et le modèle avec effets fixes. La structure du panel n'est ni totalement homogène, ni totalement hétérogène. Pour savoir si l'introduction d'effets individuels aléatoires est significativement plus pertinent que de conserver un modèle à effets fixes, nous utilisons le **test d'Hausman**. C'est un test de spécification qui permet de déterminer si les coefficients des deux estimations (fixes et aléatoires) sont statistiquement différents.

4.2 Modélisations

4.2.1 Détection de l'hétérogénéité

À partir de notre jeu de données et avant de passer aux estimations, nous pouvons d'ores et déjà constater graphiquement que les données présentent de l'hétérogénéité individuelle puisque les écarts par rapport aux valeurs moyennes par année et par pays évoluent. Nous confirmerons cela par des tests pour voir quelle estimation est la plus appropriée à notre sujet.

FIGURE 2 – Détection de l'hétérogénéité individuelle



4.2.2 Test d'homogénéité totale

Pour vérifier si statistiquement les données ne sont pas homogènes (auquel cas il ne serait pas utile d'estimer un modèle avec des effets individuels mais un seul modèle), nous allons appliquer le test de Wald à notre jeu de données. L'hypothèse nulle du test correspond à une homogénéité totale soit $H_0 : \alpha_{0i} = \alpha_0$ et $\beta' = \beta'_i \forall i$. Nous devons pour trouver la statistique, calculer la somme des carrés des résidus du modèle contraint, c'est à dire une estimation par les MCO appliquée à tous les pays à la fois. Puis nous calculerons la même statistique mais cette fois pour chaque pays séparément. Il s'agira ensuite de calculer la statistique de Fisher pour savoir s'il y a présence d'homogénéité totale dans nos données. Les résultats du test sont résumés dans la table suivante :

	Value
$SCR_{contraint}$	36,983
SCR	2,872
Test Fisher	705,378
P_value	0.000

À partir de la table ci-dessus nous rejetons l'hypothèse nulle d'homogénéité totale ce qui signifie qu'il n'est pas pertinent de réaliser une seule modélisation pour l'ensemble des pays. Il existe donc une certaine hétérogénéité dans nos données qu'il est important de prendre en compte dans les modélisations.

4.2.3 Estimations

Le test précédent nous a montré qu'il était pertinent de prendre en compte des effets individuels dans les modélisations, c'est pourquoi nous estimons 2 modèles : l'un à effet fixe et l'autre à effet aléatoire. Le tableau ci-dessous résume ces estimations avec les coefficients estimés ;

TABLE 2 – Estimations avec effets individuels

	EF - Within	EA - MCG	Within robuste
<i>Constante</i>	12.001	12.061	12.001
Taux d'emplois partiels femmes	-0.002	-0.013	-0.002
Taux chômage	0.019	0.024	0.019
Taux fécondité	-1.065	-0.866	-1.065
Taux d'anémie femmes enceintes	-0.14	-0.134	-0.14
ρ	0.899	0.738	0.899
σ_u	0.899	0.412	0.899
R^2	0.233	0.399	0.233
Nombre d'observations	176	176	176

Selon que nous considérons des effets aléatoires ou fixes, les coefficients estimés de chaque variable sont très légèrement différents, davantage pour le taux de fécondité. La qualité d'ajustement within du modèle EF est de 0.233 tandis que celle globale du modèle EA est presque 2 fois supérieure. De plus, on voit que le paramètre ρ qui donne la corrélation intraclasse est largement supérieur dans le cas d'effets fixes plutôt qu'aléatoires. Dans le premier modèle on voit que 89.9% du taux de faible poids à la naissance peut être expliqué par des différences dans les effets individuels fixes, tandis que les effets aléatoires ne peuvent expliquer que 73.8% de Y. Nous supposons alors que l'estimateur within est pour nous plus pertinent dans la modélisation de ce phénomène. Cependant, nous appliquons le test statistique de Hausman pour valider ou invalider cette hypothèse.

Le **test de Hausman** oppose H_0 : le modèle à effets aléatoires est plus approprié que celui à effets fixes, et H_1 le contraire. Étant donné que la statistique de test est de 0.001 nous rejetons l'hypothèse nulle et confirmons que le modèle *Within* est plus pertinent dans notre cas. Nous vérifions alors la validité de ce modèle final en appliquant le **test de Breusch Pagan** qui teste l'hypothèse nulle d'homoscédasticité. D'après la sortie Stata, nous voyons qu'il y a un lien entre les résidus et les variables explicatives puisque la p-value associée au test est de 0. Pour finir cette analyse, nous réestimons alors un modèle Within qui soit robuste à l'hétéroscédasticité des résidus et pouvons interpréter quelques uns des coefficients (ceux-ci étant inchangés lorsque l'on considère la version robuste ou pas).

Au seuil de risque de 10% nous voyons d'emblée que les variables du taux d'emplois partiels pour les femmes et le taux chômage ne sont pas significatives, en revanche nous pouvons tirer des conclusions des 2 autres coefficients estimés. Ainsi, lorsque le nombre d'enfants par femme augmente d'une unité, alors dans les 11 pays européens et entre 2000 et 2015, le taux d'insuffisance pondérale diminue de 1.065, ce qui peut s'expliquer par le fait que les aînés de famille sous davantage sujet à un faible poids à la naissance comme l'ont montré certaines études citées en première partie. De même, quand le taux d'anémie chez les femmes enceintes augmente d'un point de pourcentage Y diminue de 0.14 point, ce qui confirme l'importance des caractéristiques de la mère pendant la grossesse.

5 Conclusion

Pour conclure, avec les 176 observations du taux d'insuffisance pondérale expliqué par 6 facteurs explicatifs, nous réussissons entre 2000 et 2015 à expliquer 23.3% de ses variations sur les 11 pays européens de notre étude. Cependant, il est apparu lors des estimations que les variables montrant le contexte socio-économique n'étaient pas significatives dans l'explication de Y (taux de chômage et taux d'emplois partiels chez les femmes). De même, nous avons été contraints de supprimer 2 variables concernant les caractéristiques de la mère (tabagisme) et du fœtus (ratio des sexes à la naissance). La première pour cause d'un trop grand nombre de valeurs manquantes et la seconde, pour cause d'une forte corrélation avec la variable $Tx_fertile$.

Finalement, les tests ont montré qu'il existe une certaine hétérogénéité individuelle c'est pourquoi nous avons utilisé des modèles à effets fixes et aléatoires pour expliquer les différences entre pays concernant la proportion de bébés ayant un faible poids à la naissance. Il s'est alors avéré que considérer les effets fixes était plus pertinent (estimateur *within*). La vérification des hypothèses, nécessaires pour que les coefficients ne soient pas biaisés, nous a montré qu'il y avait de l'hétéroscédasticité dans les résidus, ce que nous avons résolu en estimant de manière robuste. À partir de cette dernière estimation, nous avons constaté que seules 2 de nos 6 variables de départ expliquent le faible poids des bébés à la naissance. En outre, seuls le nombre d'enfants par femmes et le taux d'anémie chez les femmes enceintes sont significatifs pour répondre à la problématique que nous nous étions fixée au début de l'analyse. De plus, nous expliquons seulement 23% de la variance totale à expliquer, ce qui nous a amené à la conclusion que nous aurions pu revoir le choix de certaines variables explicatives et notamment celles ne sortant pas significatives, dans le but d'améliorer et de compléter l'étude.